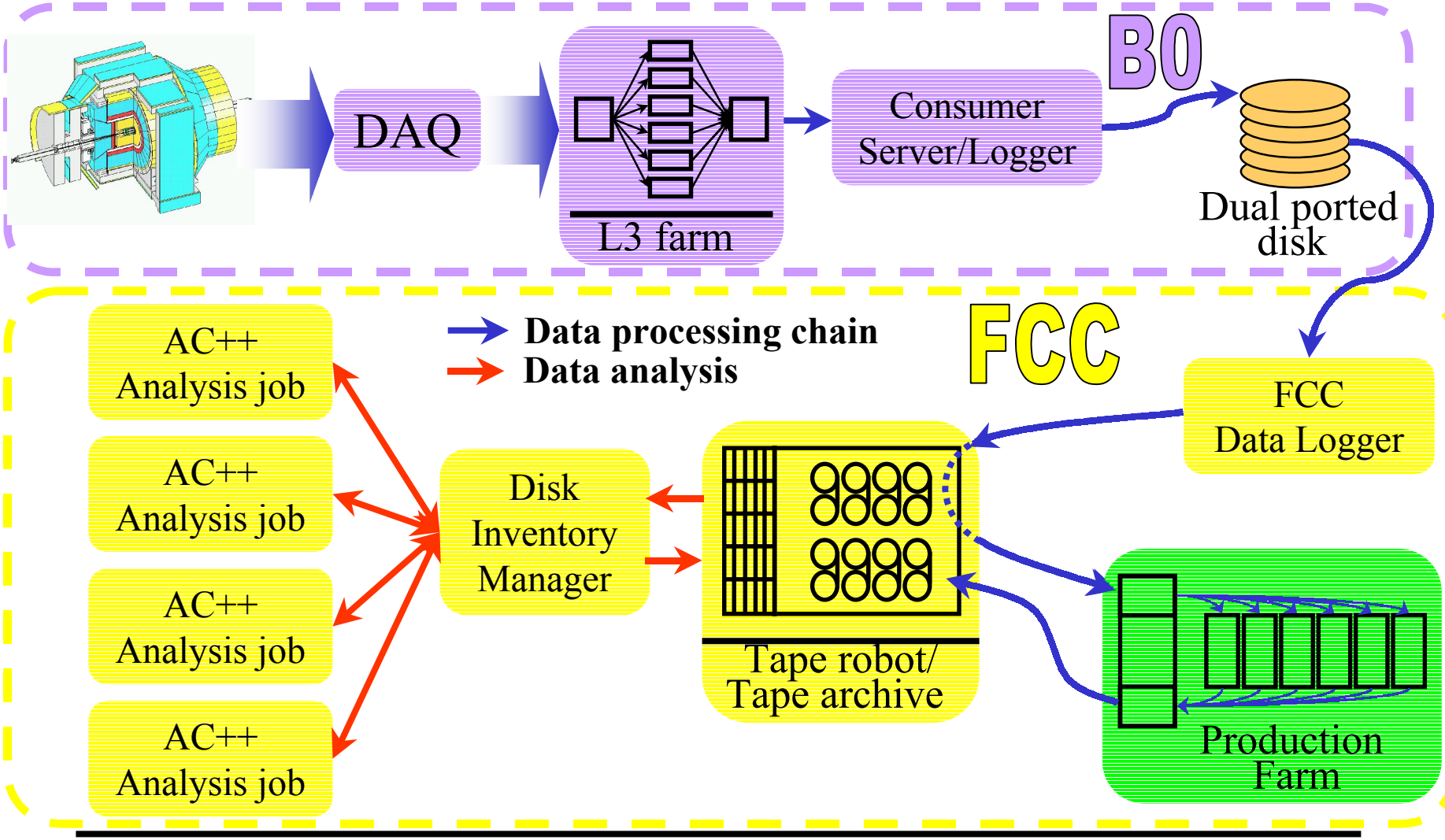
Mock Data Challenge, Part 2 Status Report

F. Ratnikov, Rutgers, **for MDC-2 team**

Yen Chu Chen, Rick St Denis, Kevin McFarland, Rob Harris, Paul Hubbard, Rob Kennedy, Stephan Lammel, Pasha Murat, Fedor Ratnikov,Liz Sexton-Kennedy, Marjorie Shapiro, Makoto Shimojama, Miroslav Siket, Rick Snider, Kirsten Tollefson, Jeff Tseng, Tony Vaiciulis, Terry Watts, Eric Wicklund, Steve Wolbers
and many others contributing to this project.

- Goals
- Components
- Schedule
- MDC-2a current status
- MDC-2b current status

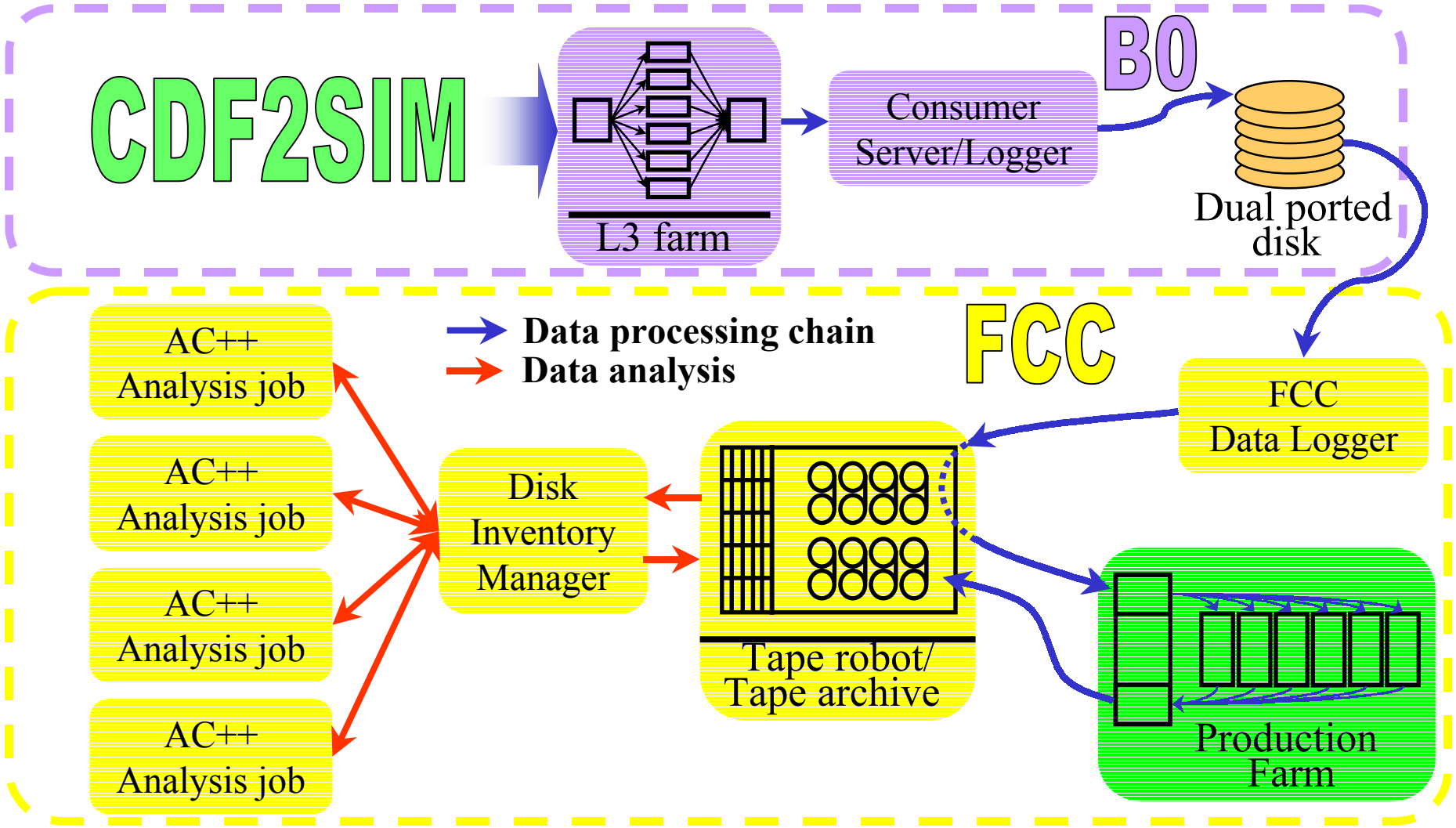
CDF Run II Data Flow



MDC Goals

- MDC-1 (end of the last century) – conductivity test
  - Data were passed through the full processing chain
  - Concepts were tested
  - Interfaces were combined together for the first time
  - Bandwidths and reliability were not part of the project
- MDC-2 (is starting now) – intensive rate test of the data processing and analysis chain
  - Test of the data transfer system at actual Run II rates
  - Exercise analysis system (L3 filters and ProductionExe) with large statistics
  - Provide collaboration with large reference sample of MC generated events
  - Store data in the Data Handling system, provide data access based on final Run II model

Mock Data Challenge



Strategy of MDC-2

Two major tasks:

- Process data chain at peak rate (20 MB/s)
- Provide large amount of data for analysis ($\approx 7 \bullet 10^6$ events)

☞ Simplify task with two stage approach:

MDC-2a: store ≈100 GB of input data locally on L3 farm, process them as many times as required, providing necessary input data rate

MDC-2b: process analysis data set separately without rate requirements

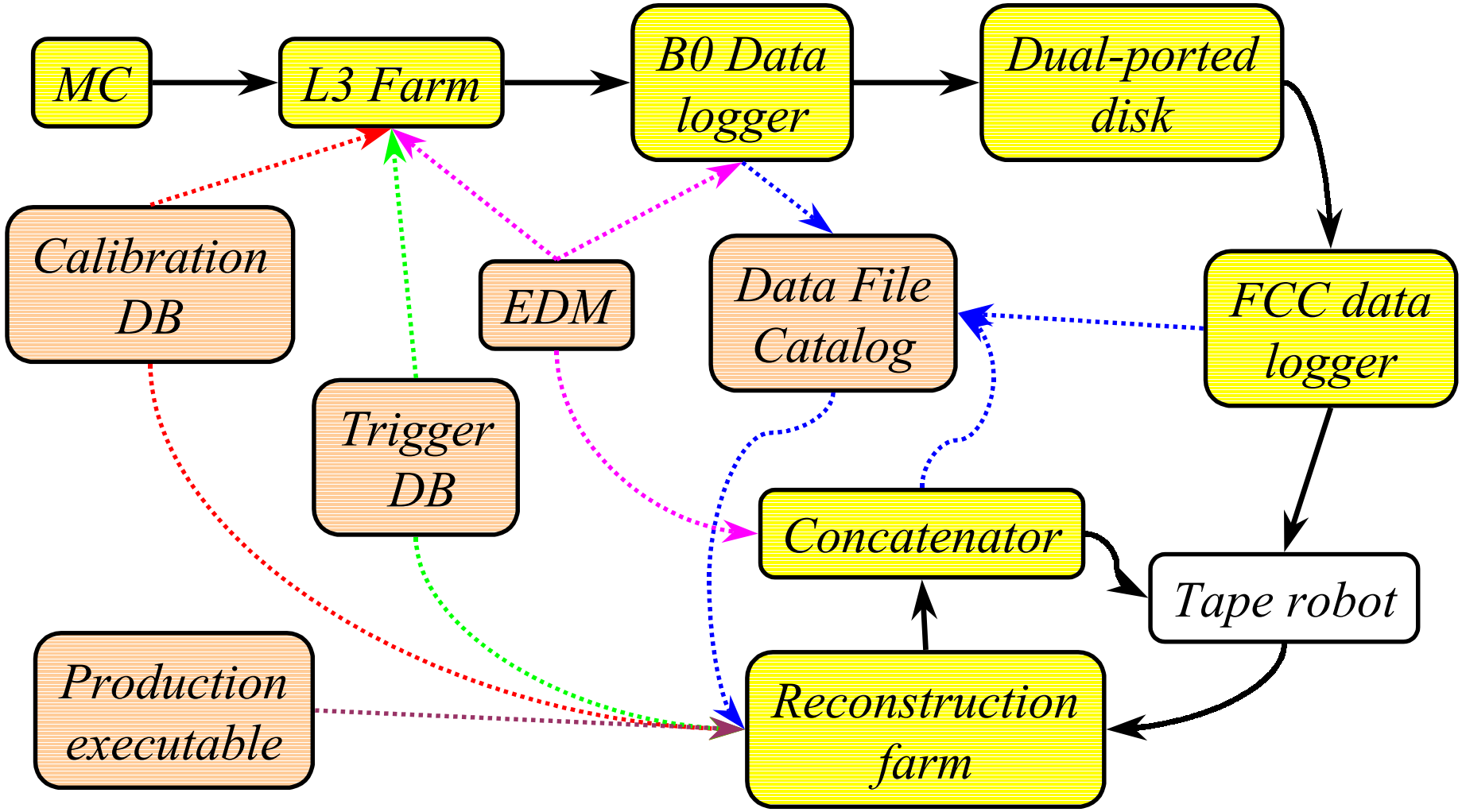
MDC-2a Strategy

- Target peak bandwidth: 20 MB/s
- 8 Sony AIT-2 tape drives are ordered
  - In accordance with approved proposal made at upgrade meeting (3/23/00) and *The Serial Media Working Group* recommendation
- 3 MB/s of write access / 6 MB/s of read access
- Full chain: 2×(tape write)+1×(tape read)

Realistic goal:

- Full chain test @6-8 MB/s
- Partial chain tests @20MB/s
  - L3→ CS/L → FCC → tape
  - Tape → production → concatenation
  - production → concatenation → tape

MDC-2a Components



MDC-2a Schedule

- Installation and testing of individual components: in progress
- 2-component / sub chain tests: May 19 $\rightarrow$ May 28
- Chain tests: May 28 $\rightarrow$ June 4
- MDC-2a rate measurements of full data chain: June 5 $\rightarrow$ June 11
- MDC-2b start: June 12

Components Status

- MC *(P. Murat)*
  - Runs on the Production Farm w/o crashes
  - Generation of 120 Gbytes of data for MDC-2a is in progress
  - To be completed and put data on local disks of L3 – end of this week
- L3 *(see talks by J.Tseng and T.Vaiciulis for details)*
  - 4 I/O + 16 processing nodes
  - Upgrade to RedHat 6.1 in progress
  - Several files processed to test system
- B0 Consumer Server/Logger (*T.Vaiciulis, M.Shimojima*)
  - New machine (b0dau32) is commissioned for CS/L
  - Output data is in Root format
  - AC++ based logger – in progress (*E. Sexton-K.*)

Components Status (cont.)

- Dual ported disk (*M. Shimojima*)
  - SGI proprietary cluster file system CXFS being installed on B0 and FCC data logger sites
- FCC Data Logger *(S. Lammel)*
  - Already used in MDC-1
  - Being expanded for full data rate
- Production Farm (*M. Siket*)
  - Runs successfully CDF2SIM
  - Generated MDC-2a input for L3
  - Generating input for MDC-2b
  - Root event I/O without unpacking objects: speedup concatenation procedure (*R. Kennedy*)

Regions of Concern

- AC++ based implementation of B0 data logger
- Concurrent access to Dual Ported disk
- Tape drives
- Speed of concatenation job on the Production farm

MDC-2b Strategy

- Physics groups were asked to specify the desirable data samples
  - Resulted: ≈7M events in total
    - $5 \bullet 10^6$ events – B-group request
    - $1.4 \bullet 10^6$ events – other groups together
- MC generation on production farm is to be completed before rate tests (in 2 weeks)
- L3 filters may run as a part of the ReconstructionExe
- Processing will take *5-10* days, starting after June, 12
- Reconstructed data samples are to be stored in the Data Handling system
- Analysis jobs will benefit from DH services

Requested Samples for MDC-2b

L1 Bit	Process	A	N(events)	Contact	event size	time/evt (sec)	size(tot) GBytes	time(tot) (hours)	L3 stream	Production stream
0	TTbar Herwig	*	100,000	Weiming	285K	12.8	28.5	355	A	T
1	Jet 20 Herwig			QCD group?					B	J
2	Jet 50 Herwig	*		QCD group?	240				B	J
3	Jet 100 Herwig			QCD group?					B	J
4	Pythia_Zee								A	T
5	Wgrad_e(+,-)	*	2x20,000	Dave Waters	193K	9.2	8.0	103	A	T
6	Wgrad_mu(+,-)	*	2x20,000	Dave Waters	190K	6.3	8.0	70	A	T
7	Wh_pythia	*	100,000	Weiming	173	~6.3	17.3	175	A	H
8	Zh_pythia	*	100,000	Weiming	170	5.4	17.0	150	A	H
9	b-->J/psi X								B	B
10	SUSY C1N2 #1	*	5,000	Jane/DaveG	150	5.0	0.75	7	A	S
11	Z -> b bbar	*	100,000	Weiming	162	5.12	16.2	150	A	H
12	Z -> tau+ tau-	*	100,000	Weiming/Pasha	120	~5.	12.0	140	A	H
13	dijets 175 Pythia	*	500,000	bjk@fnal	449	19.7	225.0	2800	B	J
14	dijets 290 Pythia	*	100,000	bjk@fnal	452	19.8	45.0	560	B	J
15	single top			Pierre					A	T
16	W+2jets(VECBOS+HERPRT)			George Velev					A	T
17	WZ	*	10,000	Dave Waters	180	6.2	1.8	17	A	T
18	WW	*	10,000	Dave Waters		-"-	1.8	17	A	T
19	ZZ	*	10,000	Dave Waters		-"-	1.8	17	A	T
20	mSUGRA C1N2->stau #1	*	25,000	Fedor/Pasha	206	6.7	5.0	46	A	S
21	mSUGRA C1N2->stau #2	*	25,000	Fedor/Pasha	239	7.8	6.0	54	A	S
22	Drell-Yan	*	50,000	Jane/DaveG	135	4.6	8.5	64	A	W
23	diphoton	*	20,000	Ray	123	4.1	2.4	23	A	
24	SUSY C1N2 #2		5,000	Jane/DaveG			0.75	7	A	S
25	SUSY C1N2 #3		5,000	Jane/DaveG			0.75	7	A	S
26	SUSY C1N2 #4		5,000	Jane/DaveG			0.75	7	A	S
27	stop production	*	50,000	Andrey Nom.	172	5.42	8.5	75	A	S
			1,400,000				415	4,894	A,B	BHJSTW
28	b bbar		5,000,000	B group	200	15-20	1,000	20,000	C	C

Requested Samples for MDC-2b

L1 Bit	Process	A	N(events)
0	TTbar Herwig	*	100,000
1	Jet 20 Herwig		
2	Jet 50 Herwig	*	
3	Jet 100 Herwig		
4	Pythia_Zee		
5	Wgrad_e(+,-)	*	2x20,000
6	Wgrad_mu(+,-)	*	2x20,000
7	Wh_pythia	*	100,000
8	Zh_pythia	*	100,000
9	b-->J/psi X		5,000
10	SUSY C1N2 #1	*	100,000
11	Z -> b bbar	*	100,000
12	Z -> tau+ tau-	*	100,000
13	dijets 175 Pythia	*	500,000
14	dijets 290 Pythia	*	100,000
15	single top		
16	W+2jets(VECBOS+HERPRT)		10,000
17	WZ	*	10,000
18	WW	*	10,000
19	ZZ	*	10,000
20	mSUGRA C1N2->stau #1	*	25,000
21	mSUGRA C1N2->stau #2	*	25,000
22	Drell-Yan	*	50,000
23	diphoton	*	20,000
24	SUSY C1N2 #2		5,000
25	SUSY C1N2 #3		5,000
26	SUSY C1N2 #4		5,000
27	stop production	*	50,000
			1,400,000
28	b bbar		5,000,000

MDC-2b Status

- All requests have been collected
  - Some requests are still a bit unclear
- Most of the corresponding .tcl files have been provided
  - production is impossible without .tcl files
- Generation of B-sample requires certain changes in the standard simulation procedure
  - We are waiting for the B-group to provide the necessary help

MC generation for MDC-2b is under way

MDC-2b Status (cont.)

- Sample files (both MC and Production) are produced for many processes and are available on fcdfsgi2 in
  - /cdf/data03/s0/mc/data/mdc2/sim
  - /cdf/data03/s0/mc/data/mdc2/prod
- Please, take a look at the process of interest to you and check that the data is suitable for your studies before generation of the full sample starts
  - That is, look today, please!
- Provide us with feedback in any case
- It will save you time in future and will also help us to use our resources effectively

Summary

- MDC-2 project is split into two parts
  - MDC-2a – test of data rate of the processing chain
  - MDC-2b – generation and reconstruction of large sample of data of various physics processes
- MDC-2a: single component test under way, chain test – last week of May
- MDC-2b: MC generation under way, reconstruction in the second half of June.